UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,550	09/15/2003	Eric Cosatto	2000-0042Con	2283

7590 05/18/2006

S. H. Dworetsky
AT&T Corp.
P.O. Box 4110
Middletown, NJ 07748

EXAMINER	
HAJNIK, DANIEL F	

ART UNIT	PAPER NUMBER
2628	

DATE MAILED: 05/18/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/662,550	COSATTO ET AL.	
	Examiner	Art Unit	
	Daniel F. Hajnik	2628	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *28 February 2006*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-35* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *22-25,27-32,34 and 35* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *15 September 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

1. Claims 22, 23, and 30 have been amended.

2. Claims 26 and 33 have been cancelled.

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 22-25, 27, 29-32, and 34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ezzat et al. (NPL document, "Visual Speech Synthesis by Morphing Visemes", herein referred to as "Ezzat") in view of Jiang et al. (NPL document, "Visual Speech Analysis with Application to Mandarin Speech Training", herein referred to as "Jiang") in view of Applicant Admitted Prior Art (herein referred to as "AAPA").

As per claims 22, 23, and 30, Ezzat teaches the claimed "selecting" step on top of 1st column on pg. 51 and states:

> "there are many intermediate frames that lie between the **chosen viseme images** ... Consequently, we compute **a series of consecutive optical flowvectors** between each intermediate image and its successor, and **concatenate** them all into one large flow vector that defines the global transformation between the chosen visemes". (emphasis added)

And states in the abstract:

> we are able to synchronize the visual speech stream with the audio speech stream, and hence give the impression of a **photorealistic talking face**. (emphasis added)

Here, the visemes represent a generic facial image that can be use to describe a particular sound and the flowvectors which contain visual and sound features are used in conjunction with the visemes.

Ezzat does not explicitly teach the claimed "obtaining" step. Jiang teaches the claimed "obtaining" step by stating in the abstract:

> **At each frame**, region of interest is identified and **key information is extracted**. The preprocessed acoustic and visual information are then fed into a modular TDNN and combined for visual speech analysis. (emphasis added)

states on (pg. 114, 4.2 Acoustic and Visual Input Representation, 1st paragraph):

> **For acoustic data representation**, we have followed the well-established approach to apply FFT on the Hamming windowed speech data to get 16 Melscale Fourier coefficients as input to the Acoustic input Layer. **For visual data representation**, we have performed the lip-tracking and feature points extraction task by applying our 2D multi-state lip shape model. Then we use both the color profile of the feature points on external and internal boundaries and position and movement of lip boundaries for feature extraction using principle component analysis (PCA). The **extracted feature vectors** are then fed to the Visual Input Layer. (emphasis added)

Here, the Jiang teaches feature vectors (target feature vector) and teaches of visual data (visual features) and acoustic information (non-visual information). It would have been obvious to one of ordinary skill in the art at the time of invention to combine Ezzat with Jiang. Jiang teaches one advantage to obtaining feature vectors in order to help children improve their speech pronunciation (see section 5, pgs. 114-115, 1st paragraph) by providing audio-visual feedback.

Ezzat does not teach the claimed "wherein generating the photo-realistic animation of the object occurs using a unit selection process". AAPA teaches the claimed limitation by teaching of "Bregler et al. utilize measurements of lip height and width, as well as teeth visibility, as visual features for unit selection" (pg. 2, lines 4-6 in the background of the submitted specification).

It would have been obvious to one of ordinary skill in the art at the time of invention to combine AAPA with the combinable system of Ezzat and Jiang. One advantage to the combination is that with AAPA unit selection features size as lip height and width measurements and teeth visibility will add a feature of more realistic facial animation generation. Further, AAPA, Ezzat, and Jiang are analogous art.

As per claims 24-25, and 31-32, Ezzat teaches the claimed "selecting ... using a comparison of a combination of visual features and non-visual features with the target feature vector" by stating on pg. 47, 2nd col, 2nd paragraph:

> For any input text, we **determine the appropriate sequence** of **viseme morphs** to make, as well as the rate of the transformations by utilizing the output of the natural language processing unit (emphasis added)

In order to determine the appropriate sequence, the system would have to perform a comparison of visual and non-visual features with a given target vector in order to produce the output as stated. Further, this construction process of an appropriate sequence of viseme morphs would require selecting candidate image samples where these samples could be used to transition between through transformation.

Ezzat teaches the claimed compiling by teaching of concatenation (see quote from top of 1st column on pg. 51 above).

As per claim 27 and 34, Ezzat teaches the claimed first database by teaching of recording and collecting one image per English phoneme (bottom of 1st column on pg. 47 under "Corpus and Viseme Acquisition", also see figure 2).

Ezzat teaches the claimed second and third database by teaching of "Flow database" (pg. 54, 2nd column), which contain optical flow vectors which specify transition data between visemes (includes visual data and includes storing non-visual data i.e. sound transitions).

As per claim 29, Ezzat teaches the claimed first database in figure 2, the claimed second database and the claimed third database on pg. 54, 2nd column under "Flow database" where this database is formed to specify visual and non-visual data between animation transitions (frames).

3. Claims 28 and 35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ezzat in view of Jiang in further view of AAPA in further of view of Brand (NPL Document, "Voice Puppetry", herein referred to as "Brand").

As per claims 28 and 35, Ezzat does not teach the claimed limitations.

Brand teaches the claimed "selecting ... a number of candidates" and the claimed "Viterbi search" by stating on the bottom half of the 1st col on pg. 25:

> The **Viterbi** sequence, while most likely, may only represent a small fraction of the total probability mass—**there may be thousands of slightly different state sequences that are nearly as likely**. If this were to happen in the voice puppet, V would be a very poor representation of the relevant information in the audio, and the animation quality would suffer greatly. ... These problems are virtually banished with entropically estimated models because **entropy minimization concentrates** the probability mass **on the optimal** Viterbi sequence. (emphasis added)

Brand teaches the claimed concatenation cost by stating on pg. 26, very bottom of 1st col and very top of 2nd col:

> We quantified this with a squared **error measure** of divergence between groundtruth (x) and reconstructed (y) facial motion vectors, **weighted to penalize motions in the wrong direction**. (emphasis added)

It would have been obvious to one of ordinary skill in the art at the time of invention to combine Brand with the combinable system of Ezzat with Jiang. Brand teaches the advantage of using an optimal Viterbi sequence with a large number of state sequences (candidates) to reduce the size to the most optimal ones in order to remove poor animation quality (1st col on pg. 25 see quote above).

Response to Arguments

4. The objection to the specification has been withdrawn in response to amendments made by applicant.

5. The objection to claim 26 has been withdrawn in response to amendments made by applicant.

6. Applicant's arguments with respect to the claims have been considered but are moot in view of the new ground(s) of rejection.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

Hon et al. (NPL Document, "AUTOMATIC GENERATION OF SYNTHESIS UNITS FOR TRAINABLE TEXT-TO-SPEECH SYSTEMS"): Section 4- "UNIT SELECTION" on pg. 295.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Daniel F. Hajnik whose telephone number is (571) 272-7642. The examiner can normally be reached on Mon-Fri (8:30A-5:00P).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ulka J. Chauhan can be reached on (571) 272-7782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

DFH Daniel Hajnik 5/3/06

ULKA CHAUHAN
SUPERVISORY PATENT EXAMINER